

## DISPLAY CONTROLLER

Patent Number: **JP5313686**  
Publication date: **1993-11-26**  
Inventor(s): **SAKURAGI KEIKO; others: 03**  
Applicant(s): **SONY CORP**  
Requested Patent:  JP5313686  
Application Number: **JP19920109357 19920402**  
Priority Number(s):  
IPC Classification: **G10L3/00**  
EC Classification:  
Equivalents:

---

### Abstract

---

PURPOSE: To output an optional voice and an image synchronously.

CONSTITUTION: A voice synthesis part 12 calculates parameters for synthesizing a voice corresponding to a text supplied from a application program 14 through a control part 11. Then, when a mouse is operated and a speak button is clicked, a control part 11 outputs an actuation instruction to a voice synthesis part 12 and an animation display part 13, and the voice synthesis part 12 outputs the parameters for synthesizing the voice corresponding to the text to a voice synthesizer, which synthesizes the voice according to the parameters. At the same time, the animation display part 13 outputs an animation corresponding to the motion of the mouse for vocalizing the text in order.

---

Data supplied from the **esp@cenet** database - I2

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平5-313686

(43)公開日 平成5年(1993)11月26日

(51)Int.Cl.  
G 10 L 3/00

識別記号 庁内整理番号  
S 8946-5H  
H 8946-5H

F I

技術表示箇所

(21)出願番号 特願平4-109357  
(22)出願日 平成4年(1992)4月2日

(71)出願人 000002185  
ソニー株式会社  
東京都品川区北品川6丁目7番35号  
(72)発明者 桜木 恵子  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内  
(72)発明者 坂口 正信  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内  
(72)発明者 浅野 薫子  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内  
(74)代理人 弁理士 稲本 義雄

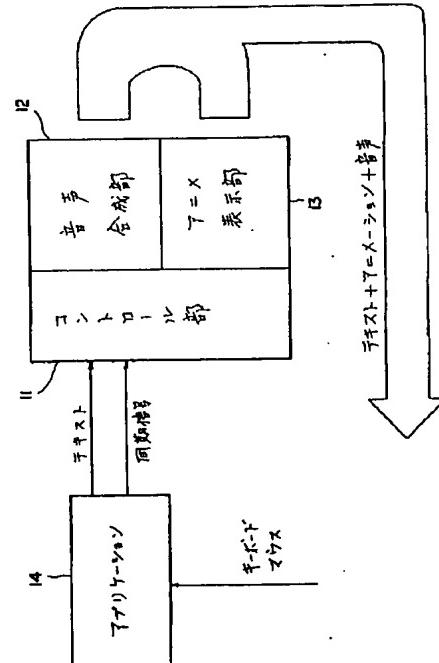
最終頁に続く

(54)【発明の名称】 表示制御装置

(57)【要約】

【目的】 任意の音声を画像と同期させて出力するこ  
ができるようとする。

【構成】 音声合成部12において、コントロール部1  
1を介してアプリケーション14より供給された、テキ  
ストに対応する音声を合成するためのパラメータが算出  
される。そして、マウス6が操作され、スピーカボタン  
35がクリックされると、コントロール部11におい  
て、起動命令が音声合成部12およびアニメ表示部13  
に出力され、音声合成部12において、テキストに対応  
する音声を合成するためのパラメータが音声合成器1  
に输出され、音声合成器1において、そのパラメータから  
音声が合成される。同時に、アニメ表示部13におい  
て、テキストを发声するための口の動きに対応するアニ  
メーションが順次出力される。



## 【特許請求の範囲】

【請求項1】 テキストから音声を合成して出力する音声合成手段と、前記テキストに対応して動画像を表示する表示手段と、前記音声合成手段により合成された音声と、前記表示手段により表示された動画像とを同期させる同期手段とを備えることを特徴とする表示制御装置。

【請求項2】 前記表示手段は、前記テキストに対応して口が動くアニメーションを表示することを特徴とする請求項1に記載の表示制御装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、例えば英語の発音を学習させるC A Iシステムなどに適用して好適な表示制御装置に関する。

## 【0002】

【従来の技術】 従来のC A Iシステムにおいては、あらかじめ、例えばハードディスクなどに記録しておいた音声信号と画像信号を読み出し、所定の同期信号に同期させて音声に対応した画像（画像に対応した音声）を出力することができるようになっている。

【0003】 これにより、例えば音声信号としてネイティブスピーカの発声した英語をハードディスクにあらかじめ記録しておくとともに、画像信号としてその英語が発声されるときの口の動き（動作）を行うアニメーションをハードディスクにあらかじめ記録しておけば、使用者に英語の発音と、その英語が発声されるときの口の動きとを同時に学習させることができる。

## 【0004】

【発明が解決しようとする課題】 このように、従来のC A Iシステムでは、あらかじめ音声信号と画像信号とをハードディスクに記録させておく必要がある。従って、例えば単語単位で記録しておいた有限数の音声信号と画像信号に対応する音声と画像しか出力することができず、使用者に学習させる範囲が限定されてしまう課題があった。

【0005】 そこで、ハードディスクに記録させておく英単語に対応する音声信号と画像信号を順次追加する方法がある。しかしながら、この方法では、ハードディスクに記録されていない音声信号と画像信号が発見されるたびに、即ち使用者が学習しているときに、学習を中断させて、その音声信号と画像信号とを追加記録しなければならないので、学習者の興味をそいでしまう課題があった。

【0006】 本発明は、このような状況に鑑みてなされたものであり、任意の音声を画像と同期させて出力することができるようとするものである。

## 【0007】

【課題を解決するための手段】 請求項1に記載の表示制御装置は、テキストから音声を合成して出力する音声合

成手段としての音声合成器1および音声合成部12と、テキストに対応して例えばアニメーションなどの動画像を表示する表示手段としてのC R T 3およびアニメ表示部13と、音声合成器1および音声合成部12により合成された音声と、C R T 3およびアニメ表示部13により表示されたアニメーションとを同期させる同期手段としてのコントロール部11とを備えることを特徴とする。

【0008】 この表示制御装置は、C R T 3およびアニメ表示部13に、テキストに対応して口が動くアニメーションを表示させることができる。

## 【0009】

【作用】 請求項1に記載の表示制御装置においては、テキストから合成された音声と、テキストに対応して動くアニメーションとを同期させて出力する。従って、任意の音声をアニメーションと同期させて出力するようすることができる。

【0010】 C R T 3およびアニメ表示部13に、テキストに対応して口が動くアニメーションを表示させることができる場合においては、例えばC A Iシステムなどに適用することにより、学習者の興味がひきつけられ、学習効果を向上させることができる。

## 【0011】

【実施例】 図1は、本発明の表示制御装置を応用したC A Iシステムの一実施例の構成を示すブロック図である。端末2は、R S 2 3 2 - C ケーブル7を介して音声合成器1に接続され、C R T 3、例えばU N I X用のワークステーション(E W S)4、キーボード5、およびマウス6より構成されている。

【0012】 E W S 4は、装置全体を制御するC P U、システムプログラムなどを記憶しているR O M、および装置の動作上必要なデータを記憶するR A M（いずれも図示せず）などより構成され、オペレーションシステムとしての例えばU N I Xオペレーションシステム、およびC R T 3に表示される（オープンされる）ウインドウなどを制御する制御プログラムとしての例えばXウインドウシステムがインストールされている。

【0013】 C R T 3は、E W S 4にインストールされたXウインドウシステムに制御され、E W S 4より出力される信号に対応して、例えばアニメーションやキーボード5で入力されたテキスト（文字列）などを表示する。キーボード5およびマウス6は、E W S 4に接続されており、テキストを作成したり、またはE W S 4に対してコマンドを発する場合に操作される。

【0014】 音声合成器1は、E W S 4よりR S 2 3 2 - C ケーブル7を介して供給される音声の特徴パラメータから音声を合成して、内蔵するスピーカ（図示せず）より出力する。

【0015】 図1に示すC A Iシステムは、英語学習用のもので、学習者が自由に試行錯誤しながら能動的に思

考することができる環境提供型、且つコミュニケーション型（双方主導型）のシステムであり、

○シミュレーション機能

○検索機能

○ユーザ・フレンドリなインターフェース

を有し、マルチメディアを駆使して、範囲にとらわれない学習、および様々な学習方法を提供することができるようになっている。

【0016】さらに、このCAIシステムは、楽しい学校の雰囲気を出すため、

- ・「STUDY」
- ・「POST OFFICE」
- ・「LIBRARY」
- ・「PLAYROOM」
- ・「LABORATORY」

と呼ばれる5つのサブシステムから構成される。

【0017】サブシステム「STUDY」は、教科書に基づいて英語の文法的基礎を学び、内容を理解し、さらには知識の定着をはかるための問題演習などを行なう、基礎学習室である。

【0018】サブシステム「POST OFFICE」は、実用的な英語での表現や手紙の書式を身につける手助けをするために、英文で手紙を書いて送ることをシミュレートするものである。

【0019】サブシステム「LIBRARY」は、システム（各サブシステム）内で使用される言葉をデータベースとして持ち、他のサブシステムから自由に検索・参照できる支援ツールである。このサブシステムでは、絵や音も提供されるので、日本語を経由することなく視覚聴覚の助けによって、学習者に言葉のイメージを捕えさせることができるようになっている。

【0020】サブシステム「PLAYROOM」は、英語の勉強に対する、学習者の興味、関心が失われてしまうのを防止するためのもので、知的なゲームで遊びながら、知らず知らずのうちに、英語を学習することができるようになっている。

【0021】サブシステム「LABORATORY」は、英語による会話の実験をするためのもので、画像と音声を組合せ、実際に人と会話しているかのような状況（仮想状況）を作り、その中で学習者に英会話を体験させることができるようになっている。

【0022】サブシステム「LABORATORY」は、アプリケーション14、並びに実際に人と会話しているかのような状況（仮想状況）を作り出すためのユーザインターフェースとしてのコントロール部11、音声合成部12、およびアニメ表示部13より構成される（図2）。

【0023】コントロール部11は、音声合成部12およびアニメ表示部13を制御し、音声合成部12（音声合成器1）により合成される音声と、アニメ表示部13

によりCRT3に表示されるアニメーションの口の動きとを同期させる。さらに、コントロール部11は、アプリケーション14より供給されるテキストの文字数、テキストに含まれるカンマ(,)の数、またはビリオド(.)の数をそれぞれカウントし、CRT3に表示されるアニメーションの表示時間（アニメーションの口を動かす時間）を算出してアニメ表示部13に出力する。音声合成部12は、コントロール部11を介してアプリケーション14より供給されるテキスト（文字列）から、音声を合成するためのパラメータを算出し、音声合成器1に出力する。アニメ表示部13は、コントロール部11の制御にしたがって、セルフレームデータ記憶部13a（図6）に記憶された、アニメーション（アニメーションの口）を動かすためのセルフレームデータを読み出し、CRT3に出力する。セルフレームデータ記憶部13aには、アニメーション（アニメーションの口）を動かすためのセルフレームデータ（図8）が記憶されている。

【0024】アプリケーション14は、サブシステム「LABORATORY」におけるアプリケーションで、例えば図3に示すように、対話処理部21、言語処理部22、知識データベース検索部23、および知識データベース24より構成される。対話処理部21は、様々な状況に関するスクリプトと対話履歴を記憶（保持）し、これらを基に対話の流れを管理する。

【0025】言語処理部22は、解析部と生成部（いずれも図示せず）より構成される。言語処理部22の解析部は、キーボード5で入力された英語文の構文解析、および意味解析を行い、内部表現に変換して、知識データベース検索部23に供給する。言語処理部22の生成部は、知識データベース検索部23より返された回答と、対話処理部21からの指示（制御）に基づき英語文を生成し、コントロール部11（図2）に出力する。

【0026】知識データベース検索部23は、言語処理部22の解析部より供給された内部表現に対する返答を、知識データベース24に記憶されているデータベースから検索して言語処理部22の生成部に出力する。知識データベース24には、質問に対する返答パターンなどが内部表現の形で記憶されている。

【0027】次に、その動作について説明する。まず、図1に示すCAIシステムにおいて、キーボード5またはマウス6が操作され、図3に示すアプリケーション14が起動されると、CRT3に図4に示すような、先生としてのアニメーション32と、CAIシステムの使用者、即ち学習者本人としてのアニメーション31が表示される。さらに、キーボード5が操作され、質問文としてのテキスト33（What is the fastest train in Japan?）が入力されると、Xウインドウシステムに制御されたCRT3のウインドウ上にテキスト33が表示され

るとともに、アプリケーション14の言語処理部22

(図3)にテキスト33が供給される。

【0028】言語処理部22の解析部において、キーボード5で入力されたテキスト33の構文解析、および意味解析が行われ、内部表現に変換されて、知識データベース検索部23に供給される。知識データベース検索部23において、言語処理部22の解析部より供給された内部表現に対する返答が、知識データベース24に記憶されているデータベースから検索され言語処理部22の生成部に出力される。言語処理部22の生成部において、知識データベース検索部23より返された回答と、対話処理部21からの指示(制御)に基づきテキスト34(図4)(The fastest train in Japan is Shinkansen.)が生成され、コントロール部11(図2)に出力される。

【0029】コントロール部11において、アプリケーション14より供給されたテキスト34の文字数、テキスト34に含まれるカンマ(,)の数、またはビリオド(.)の数がそれぞれカウントされ、CRT3に表示された先生としてのアニメーション32の表示時間(アニメーション32の口を動かす時間)が算出され、アニメ表示部13に供給される。

【0030】アニメ表示部13において、コントロール部11を介してアプリケーション14より供給された、テキスト33の返答であるテキスト34を発声するための口の動きに対応するアニメーション32のセルフレームデータが、セルフレームデータ記憶部13aより読み出される。

【0031】一方、音声合成部12において、コントロール部11を介してアプリケーション14より供給された、テキスト33の返答であるテキスト34に対応する音声を合成するためのパラメータが算出される。

【0032】そして、マウス6が操作され、スピーカボタン35がクリックされると、コントロール部11において、起動命令が音声合成部12およびアニメ表示部3に出力される。

【0033】音声合成部12において、テキスト34に対応する音声を合成するためのパラメータが音声合成器1に出力され、音声合成器1において、そのパラメータから音声「The fastest train in Japan is Shinkansen.」が合成され、内蔵するスピーカより出力される(図5)。

【0034】同時に、アニメ表示部13において、セルフレームデータ記憶部13aより読み出された、テキスト34を発声するための口の動きに対応するアニメーション32のセルフレームデータが、コントロール部11より供給されたアニメーション32の表示時間(アニメーション32の口を動かす時間)に対応して、CRT3に順次出力され(図6)、CRT3において、テキスト34を発声しているように口を動かすアニメーション32が表示される。

【0035】なお、このとき、Xウインドウシステムに制御されたCRT3のウインドウ上にテキスト34(図4)が表示される。

【0036】以上のようにして、テキストから合成された音声と、テキストに対応して動くアニメーションとが同期して出力される。

【0037】次に、図7を参照して、さらにその動作について説明する。まず最初に、ステップS1において、キーボード5よりテキストが入力されたか否かが判定される。ステップS1において、キーボード5よりテキストが入力されていないと判定された場合、再びステップS1に戻る。ステップS1において、キーボード5よりテキストが入力されたと判定された場合、ステップS2に進み、テキストに対応する口の動きを行うためのアニメーションのセルフレームデータがセルフレームデータ記憶部13aから読み出されるとともに、そのセルフレームデータをCRT3に出力する時間、即ちアニメーションの口を動かす時間が算出され、ステップS3に進む。

【0038】ステップS3において、ステップS1で入力されたテキストから音声合成するためのパラメータが算出され、ステップS4に進む。ステップS4において、ステップS3で算出されたパラメータから合成される音声の出力タイミングと、ステップS2でセルフレームデータ記憶部13aから読み出されたセルフレームデータの出力タイミングとを同期させるための同期信号(起動命令)が検出されたか否かが判定される。ステップS4において、同期信号が検出されていないと判定された場合、再びステップS4に戻る。ステップS4において、同期信号が検出されたと判定された場合、ステップS5に進み、ステップS3で算出されたパラメータから合成された音声が、音声合成器1の内蔵するスピーカより出力されるとともに、ステップS1で入力されたテキストを発声しているように口を動かすアニメーションが、ステップS2で算出されたアニメーションの口を動かす時間に対応して、CRT3に表示される。

【0039】次に、アニメーションの表示時間(アニメーションの口を動かす時間)の算出方法について説明する。まず、セルフレームデータ記憶部13aに記憶されているセルフレーム(図8)をアニメーションらしく表示させるために、CRT3で一枚のセルフレームが描画される回数をTとする。なお、この回数Tは、CRT3の表示速度と画像データの大きさを基にして、あらかじめEWS4に設定されている。

【0040】CRT3に表示するセルフレームの枚数をNとすると、合成された音声が出力されている時間内に、時間(T×N)だけかかるセルフレームの表示を行うことのできる回数TIMESは、テキストの文字数をMとして、次式にしたがって求められる。

$$50 \text{ TIMES} = M/a$$

但し、 $a$ はテキスト中にコンマ(,)、またはビリオド(.)を含むか否かと、これらを含む場合はその数によって可変の値を取るようになっており、取りうる値は前述したTと同様、あらかじめEWS4に設定されている。

【0041】ここで、図9はテキストの文字数と発話時間の関係を表す図である。テキストに用いられている単語やその前後との繋がりにより若干ばらつきがあるものの、図に示すとおりほぼ矩形領域内に収まる。(この統計結果によると、1文字あたりの発話時間は、0.05秒程度となっている。)

【0042】上述した $a$ が可変値を取る理由は、例えばテキスト「No. Japan is not a large country.」や、テキスト「Yes, I do.」などのようにテキスト中にビリオド、またはカンマがある場合、発話においては、その後に一呼吸、間が入るので、全体としての発話時間が若干長くなることを考慮しているためである。

【0043】図10は、音声合成器1からのテキストの発話時間を測定した結果を示す図である。図10(a)はテキスト「No. Japan is not a large country.」の波形で、図10(b)はテキスト「No Japan is not a large country.」の波形で、図10(c)はテキスト「Now Japan is not a large country.」の波形である。この3つを比べると、ビリオドがテキスト中に入っているテキスト「No. Japan is not a large country.」の発話時間が一番長いことが判る。

【0044】以上説明した方法により、テキストから合成された音声と、そのテキストに対応して動くアニメーションの同期が、視聴覚的許容範囲内で実現可能となる。

【0045】さらに、音声合成器1が音声出力を終えた時点で、終了信号をEWS4に返すようにして、EWS4において、この終了信号が検出された場合に動画の表示を終了するようにすれば、テキストから合成された音声と、そのテキストに対応して動くアニメーションを、より良く同期させることができる。

【0046】また、単語の発話時間が定義してある単語辞書(図11)をEWS4の内蔵する、例えばハードディスク(図示せず)にあらかじめ記録させておくようすれば、例えばテキスト「My uncle lives in a small town.」などのテキストが与えられた時点で、そのテキストを構成する単語「my」、「uncle」、「live」、「in」、「a」、「small」、「town」の発話時間を単語辞書より、それぞれ0.148秒、0.298秒、0.275秒、0.103秒、0.056秒、0.319秒、0.231秒のように算出し、この合計時間(1.928秒)の間だけアニメーションを表示するようになるとができる。

【0047】さらに、図8に示す口の動きを表すセルフレームと、英語の発音記号との対応表(図12)をEWS4

S4の内蔵するハードディスクにあらかじめ記録させておくようすれば、テキストを発音記号に変換し(図13)、対応表から表示するセルフレームを得るようにして、アニメーションを生成するようにすることができる。

【0048】以上説明したように、テキストから合成された音声が、テキストに対応して動くアニメーションが話しているかのように出力されるので、学習者の興味がひきつけられ、学習効果を向上させることができる。

【0049】

【発明の効果】請求項1に記載の表示制御装置によれば、テキストから合成された音声と、テキストに対応して表示される動画とが同期して出力される。従って、任意の音声を動画と同期させて出力するようにすることができる。

【0050】請求項2に記載の表示制御装置によれば、表示手段に、テキストに対応して口が動くアニメーションを表示させるようにしたので、例えばCAIシステムなどの学習用装置に適用することにより、学習者の興味がひきつけられ、学習効果を向上させることができる。

【図面の簡単な説明】

【図1】本発明の表示制御装置を応用したCAIシステムの一実施例の構成を示すブロック図である。

【図2】図1のCAIシステムのサブシステム「LABORATORY」におけるユーザインターフェースを示す図である。

【図3】図2のアプリケーション14のより詳細なブロック図である。

【図4】図1のCAIシステムのCRT3に表示される画面を示す図である。

【図5】図2の音声合成部12の入出力を説明するための図である。

【図6】図2のアニメ表示部13の入出力を説明するための図である。

【図7】図2のインターフェースの動作を説明するためのフローチャートである。

【図8】アニメ表示部13のセルフレームデータ記憶部13aに記憶されているセルフレームデータを示す図である。

【図9】テキストの文字数と、テキストの発声時間との関係を測定した結果を示す図である。

【図10】カンマまたはビリオドを含むテキストと、含まないテキストとの発話時間を示す図である。

【図11】単語と、単語の発話時間とを対応させた単語辞書を示す図である。

【図12】図8のセルフレームデータと、発音記号との対応表を示す図である。

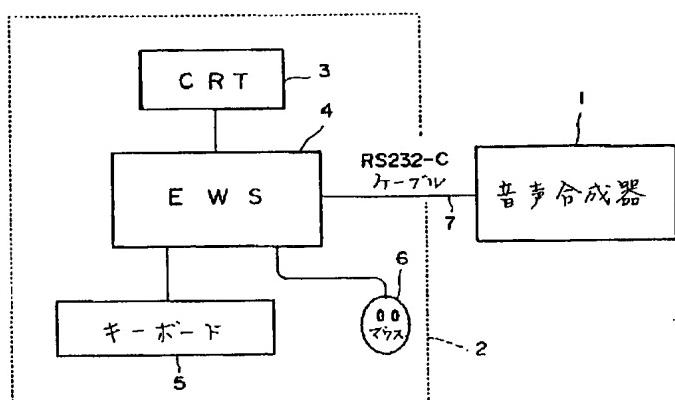
【図13】テキストを発音記号に変換したことを示す図である。

50 【符号の説明】

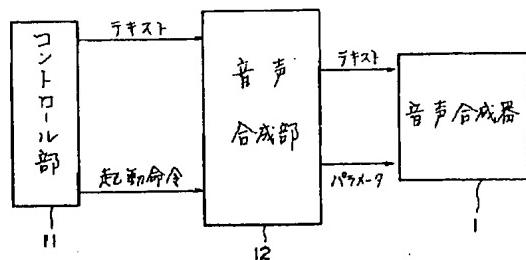
- 9
- 1 音声合成器
  - 2 端末
  - 3 CRT
  - 4 ワークステーション (EWS)
  - 5 キーボード
  - 6 マウス
  - 11 コントロール部
  - 12 音声合成部
  - 13 アニメ表示部

- \* 13 a セルフレームデータ記憶部  
 14 アプリケーション  
 21 対話処理部  
 22 言語処理部  
 23 知識データベース検索部  
 24 知識データベース  
 31, 32 アニメーション  
 33, 34 テキスト  
 \* 35 スピーカボタン

【図1】

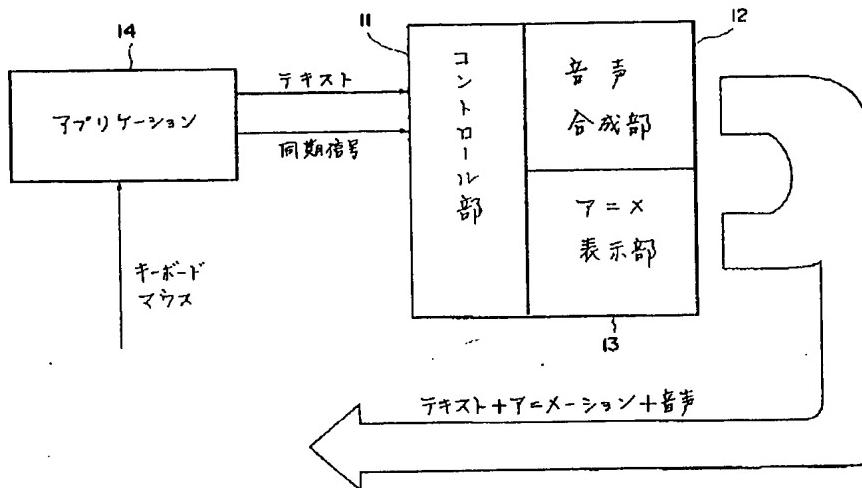


【図5】



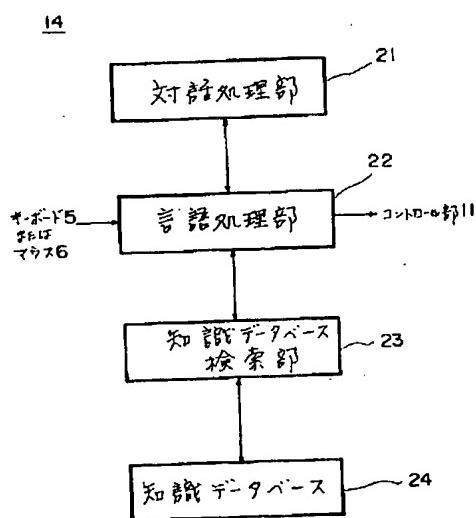
【図11】

【図2】

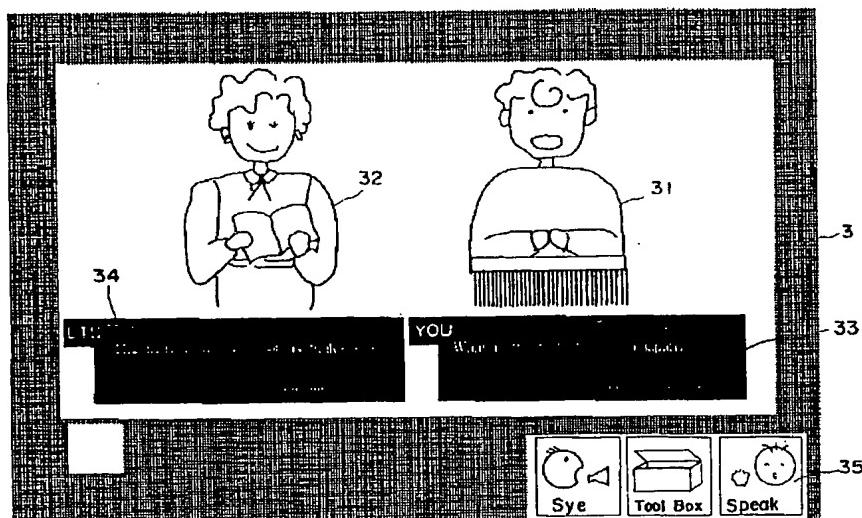


単語	発話時間 (秒)
go	0.187
live	0.275
uncle	0.298
a	0.056
in	0.103
house	0.311
town	0.231
small	0.319
my	0.148

【図3】

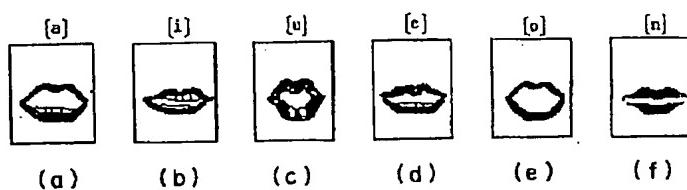


【図4】



発音記号	セリフレーム
(e)	[e]
(ʌ)	[ə]
(i), (i:)	[ɪ]
(u), (u:)	[ʊ]
(əi)	[ə]+[ɪ]
(əə)	[ə]
(əu)	[ə]+[u]
(ŋ)	[ŋ]

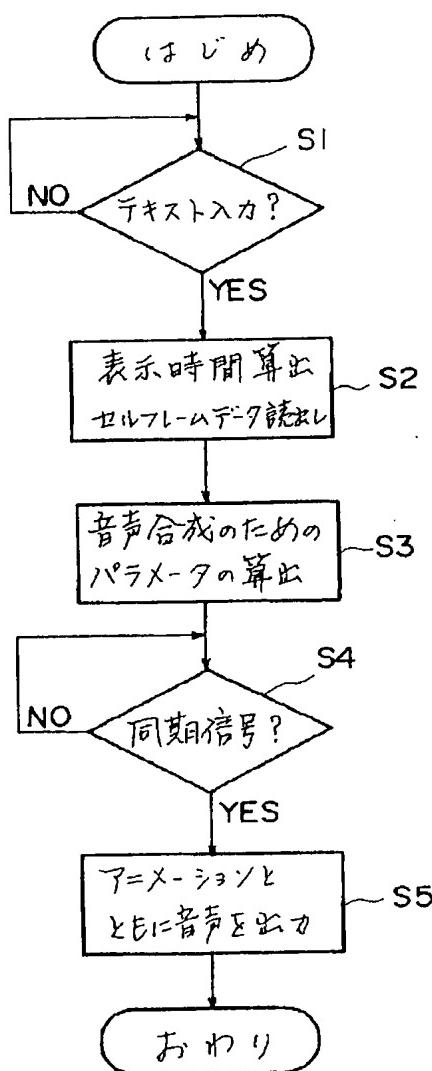
【図8】



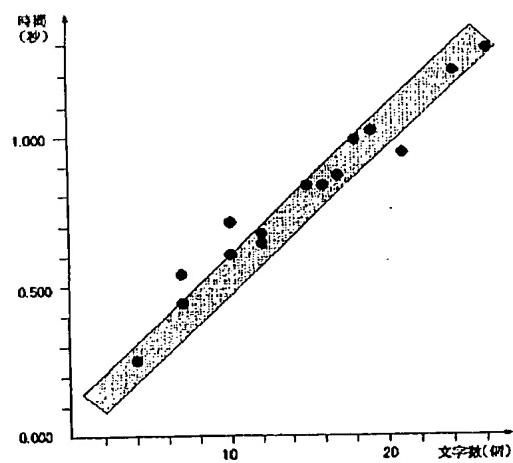
【図13】

(a) I have a pencil.  
 (b) (əi) (hæv) (ə) (pɛnsl)

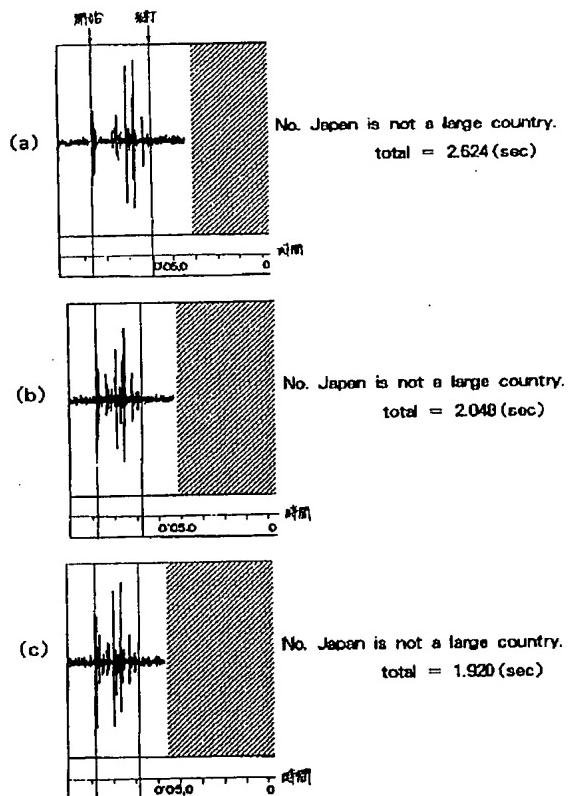
【図7】



【図9】



【図10】



フロントページの続き

(72)発明者 川手 史隆  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第2区分

【発行日】平成11年(1999)11月26日

【公開番号】特開平5-313686

【公開日】平成5年(1993)11月26日

【年通号数】公開特許公報5-3137

【出願番号】特願平4-109357

【国際特許分類第6版】

G10L 3/00

【F I】

G10L 3/00

S

H

【手続補正書】

【提出日】平成11年4月2日

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正内容】

【特許請求の範囲】

【請求項1】 入力される文字列から音声を合成して出力する音声合成手段と、

前記文字列に関する諸データと、前記文字列に対応した動画像表示のための画像情報群を有する記憶手段と、

前記文字列に対応した動画像を表示するために、前記記憶手段から画像情報群を読み出して表示する動画像表示手段と、

前記文字列から得られる制御データと前記記憶手段の諸データに基づいて、前記音声合成手段により合成された音声と、前記動画像表示手段により表示された動画像とを同期させる同期制御手段とを備えることを特徴とする表示制御装置。

【請求項2】 前記記憶手段に記憶される前記文字列に関する諸データは、前記音声合成手段が合成音声を出力するのに要する各単語毎の出力時間情報、または各単語毎の発音記号情報であり、

前記記憶手段に記憶される画像情報群は、前記文字列に対応した動画像を表示させるための静止画群であり、

前記同期制御手段は、前記文字列から得られる制御データおよび前記記憶手段の諸データに基づいて、前記音声合成手段と前記動画像表示手段とに動作開始命令を与えるとともに、前記音声合成手段からの音声出力終了に同期するように前記動画像表示手段に動作終了命令を与えることを特徴とする請求項1に記載の表示制御装置。

【請求項3】 前記文字列は英文テキストであり、

前記文字列に対応した動画像は、前記英文テキストを発話している様子を示すアニメーションであることを特徴

とする請求項2に記載の表示制御装置。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0007

【補正方法】変更

【補正内容】

【0007】

【課題を解決するための手段】本発明の表示制御装置は、入力される文字列から音声を合成して出力する音声合成手段と、文字列に関する諸データと、文字列に対応した動画像表示のための画像情報群を有する記憶手段と、文字列に対応した動画像を表示するために、記憶手段から画像情報群を読み出して表示する動画像表示手段と、文字列から得られる制御データと記憶手段の諸データに基づいて、音声合成手段により合成された音声と、動画像表示手段により表示された動画像とを同期させる同期制御手段とを備えることを特徴とする。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0008

【補正方法】削除

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0009

【補正方法】変更

【補正内容】

【0009】

【作用】上記構成の表示制御装置においては、入力文列から合成された音声と、その入力文列に対応した動画像とを同期させて出力する。従って、任意の音声を動画像と同期させて出力するようになることができる。

【手續補正5】

【補正対象書類名】明細書

【補正対象項目名】0010

【補正方法】削除

【手続補正6】

【補正対象書類名】明細書

【補正対象項目名】0011

【補正方法】変更

【補正内容】

【0011】

【実施例】図1は、本発明の表示制御装置を応用したCAIシステムの一実施例の構成を示すブロック図である。端末2は、RS232-Cケーブル7を介して音声合成器1に接続され、CRT3、例えばUNIX用のワークステーション(EWS)などのコンピュータ4、キーボード5、およびマウス6より構成されている。

【手続補正7】

【補正対象書類名】明細書

【補正対象項目名】0012

【補正方法】変更

【補正内容】

【0012】コンピュータ4は、装置全体を制御するCPU、システムプログラムなどを記憶しているROM、および装置の動作上必要なデータを記憶するRAM(いずれも図示せず)などより構成され、オペレーティングシステムとしての例えばUNIXオペレーティングシステム、およびCRT3に表示される(オープンされる)ウインドウなどを制御する制御プログラムとしての例えばXウインドウシステムがインストールされている。本実施例においては、コンピュータ4はEWS、オペレーティングシステムはUNIXとしたが、コンピュータ4はパーソナルコンピュータ(PC)でもよく、オペレーティングシステムもUNIXに限らない。

【手続補正8】

【補正対象書類名】明細書

【補正対象項目名】0013

【補正方法】変更

【補正内容】

【0013】CRT3は、コンピュータ4にインストールされたXウインドウシステムに制御され、コンピュータ4より出力される信号に対応して、例えばアニメーションやキーボード5で入力されたテキスト(文字列)などを表示する。キーボード5およびマウス6は、コンピュータ4に接続されており、テキストを作成したり、またはコンピュータ4に対してコマンドを発する場合に操作される。

【手続補正9】

【補正対象書類名】明細書

【補正対象項目名】0014

【補正方法】変更

【補正内容】

【0014】音声合成器1は、コンピュータ4よりRS232-Cケーブル7を介して供給されるテキスト(文

字列)とそれに対応する音声の特徴パラメータとから音声を合成して、内蔵するスピーカ(図示せず)より出力する。この音声合成器1は、コンピュータに挿入するタイプの音声合成ボードでもよく、本実施例においては、外部接続タイプのDEC社製音声合成器を採用した。これは、入力がテキスト(文字列)およびその関係データで、出力が合成音声である機能を備える装置である。

【手続補正10】

【補正対象書類名】明細書

【補正対象項目名】0015

【補正方法】変更

【補正内容】

【0015】図1に示すCAIシステムは、英語学習用のもので、学習者が自由に試行錯誤しながら能動的に思考することができる環境提供型、且つコミュニケーション型(双方主導型)の学習システムであり、

○シミュレーション機能

○検索機能

○ユーザ・フレンドリなインターフェース

を有し、マルチメディアを駆使して、範囲にとらわれない学習、および様々な学習方法を提供することができるようになっている。

【手続補正11】

【補正対象書類名】明細書

【補正対象項目名】0020

【補正方法】変更

【補正内容】

【0020】サブシステム「PLAYROOM」は、英語の勉強に対する、学習者の興味、関心が失われてしまうのを防止するためのもので、知的なゲームで遊びながら、知らず知らずのうちに、英語を学習することができるよう~~に工夫がなされている~~。

【手続補正12】

【補正対象書類名】明細書

【補正対象項目名】0021

【補正方法】変更

【補正内容】

【0021】サブシステム「LABORATORY」は、英語による会話の実験をするためのもので、画像と音声を組合せ、実際に人と会話しているかのような状況(仮想状況)を作り、その中で学習者に擬似的に英会話を体験させることができるようになっている。

【手続補正13】

【補正対象書類名】明細書

【補正対象項目名】0022

【補正方法】変更

【補正内容】

【0022】サブシステム「LABORATORY」は、アプリケーション14、並びに実際に人と会話しているかのような状況(仮想状況)を作り出すためのユー

ザインターフェース部からなり、インターフェース部は、コントロール部11、音声合成部12、およびアニメ表示部13より構成される(図2)。

## 【手続補正14】

【補正対象書類名】明細書  
【補正対象項目名】0023

## 【補正方法】変更

## 【補正内容】

【0023】コントロール部11は、音声合成部12およびアニメ表示部13を制御し、音声合成部12(音声合成器1)により合成される音声と、アニメ表示部13によりCRT3に表示されるアニメーションの口の動きとを同期させる。さらに、コントロール部11は、アプリケーション14より供給されるテキストの文字数、テキストに含まれるカンマ(,)の数、またはビリオド(.)の数をそれぞれカウントし、CRT3に表示されるアニメーションの表示時間(アニメーションの口を動かす時間)を算出してアニメ表示部13に出力する。音声合成部12は、コントロール部11を介してアプリケーション14より供給されるテキスト(文字列)から、音声を合成するためのパラメータを算出し、音声合成器1に出力する。アニメ表示部13は、コントロール部11の制御にしたがって、セルフレームデータ記憶部13a(図6)に記憶された、アニメーションを表示させるためのセルフレームデータを読み出し、CRT3に出力する。セルフレームデータ記憶部13aには、アニメーション(英語をしゃべっている様子のアニメーション)を表示させるためのセルフレームデータ(図8)が記憶されている。

## 【手続補正15】

【補正対象書類名】明細書  
【補正対象項目名】0030

## 【補正方法】変更

## 【補正内容】

【0030】次に、コントロール部11を介してアプリケーション14より供給されたテキスト33の返答であるテキスト34を発声するための口の動きに対応するアニメーション32のセルフレームデータが、アニメ表示部13においてセルフレームデータ記憶部13aより読み出される。

## 【手続補正16】

【補正対象書類名】明細書  
【補正対象項目名】0033

## 【補正方法】変更

## 【補正内容】

【0033】音声合成部12において、テキスト34に対応する音声を合成するためのパラメータが音声合成器1に出力され、音声合成器1において、そのパラメータに基づいて音声「The fastest train in Japan is Shinkansen.」が合成され、内蔵するスピーカより出力され

る(図5)。

## 【手続補正17】

【補正対象書類名】明細書

【補正対象項目名】0037

## 【補正方法】変更

## 【補正内容】

【0037】次に、図7を参照して、さらにその動作について説明する。まず最初に、ステップS1において、キーボード5よりテキストが入力されたか否かが判定される。ステップS1において、キーボード5よりテキストが入力されていないと判定された場合、再びステップS1に戻る。ステップS1において、キーボード5よりテキストが入力されたと判定された場合、ステップS2に進み、テキストに対応する口の動きを表現するためのアニメーションのセルフレームデータがセルフレームデータ記憶部13aから読み出されるとともに、そのセルフレームデータをCRT3に出力する時間、即ちアニメーションの口を動かす時間が算出され、ステップS3に進む。

## 【手続補正18】

【補正対象書類名】明細書

【補正対象項目名】0039

## 【補正方法】変更

## 【補正内容】

【0039】次に、アニメーションの表示時間(アニメーションの口を動かす時間)の算出方法について説明する。まず、アニメーションの基本となる動作を構成するためのセルフレームN枚を、セルフレームデータ記憶部13aに登録する。アニメーション表示は、この基本となる動作表示を、合成された音声が出力されている時間内、繰り返し行うことによって実現する。

## 【手続補正19】

【補正対象書類名】明細書

【補正対象項目名】0040

## 【補正方法】変更

## 【補正内容】

【0040】合成された音声の出力時間TIME(理論値)は、実際の音声合成器の1文字あたりの平均出力時間aと、テキストの文字数Mとから、次式にしたがって求める。

$$TIME = a \times M \quad (\text{sec})$$

このTIME時間内、セルフレームを繰り返し表示させ、時間経過と同時に停止されることによって、音声と同期したアニメーション表示を実現させている。但し、aは、テキスト中にカンマ(,)、ビリオド(.)を含むか否かと、これらを含む場合はその数によって可変の値を取るようになっている。

## 【手続補正20】

【補正対象書類名】明細書

【補正対象項目名】0045

【補正方法】変更

【補正内容】

【0045】さらに、音声合成器1が音声出力を終えた時点で、終了信号をコンピュータ4に返すようにして、コンピュータ4において、この終了信号が検出された場合に動画の表示を終了するようにすれば、テキストから合成された音声と、そのテキストに対応して動くアニメーションを、より良く同期させることができる。

【手続補正21】

【補正対象書類名】明細書

【補正対象項目名】0046

【補正方法】変更

【補正内容】

【0046】また、単語の発話時間が定義してある単語辞書(図11)をコンピュータ4の内蔵する、例えばハードディスク(図示せず)にあらかじめ記録させておくようすれば、例えばテキスト「My uncle lives in a small town.」などのテキストが与えられた時点で、そのテキストを構成する単語「my」、「uncle」、「live」、「in」、「a」、「small」、「town」の発話時間を単語辞書より、それぞれ0.148秒、0.298秒、0.275秒、0.103秒、0.056秒、0.319秒、0.231秒のように算出し、この合計時間(1.928秒)の間だけアニメーションを表示するようになることができる。

【手続補正22】

【補正対象書類名】明細書

【補正対象項目名】0047

【補正方法】変更

【補正内容】

【0047】さらに、図8に示す口の動きを表すセルフレームと、英語の発音記号との対応表(図12)をコンピュータ4の内蔵するハードディスクにあらかじめ記録させておくようすれば、テキストを発音記号に変換し(図13)、対応表から表示するセルフレームを得るようにして、アニメーションを生成するようにすることができる。

【手続補正23】

【補正対象書類名】明細書

【補正対象項目名】0048

【補正方法】変更

【補正内容】

【0048】以上説明したように、テキストから合成された音声が、テキストに対応してアニメーションが話しているかのように出力されるので、学習者の興味がひきつけられ、学習効果を向上させることができる。

【手続補正24】

【補正対象書類名】明細書

【補正対象項目名】0049

【補正方法】変更

【補正内容】

【0049】

【発明の効果】以上如く、本発明の表示制御装置によれば、文字列から合成された音声と、その文字列テキストに対応して表示される動画像とが同期して出力される。従って、任意の音声を動画像と同期させて出力するようになることができる。

【手続補正25】

【補正対象書類名】明細書

【補正対象項目名】0050

【補正方法】削除

【手続補正26】

【補正対象書類名】明細書

【補正対象項目名】符号の説明

【補正方法】変更

【補正内容】

【符号の説明】

1 音声合成器

2 端末

3 CRT

4 コンピュータ

5 キーボード

6 マウス

1 1 コントロール部

1 2 音声合成部

1 3 アニメ表示部

1 3 a セルフレームデータ記憶部

1 4 アプリケーション

2 1 対話処理部

2 2 言語処理部

2 3 知識データベース検索部

2 4 知識データベース

3 1, 3 2 アニメーション

3 3, 3 4 テキスト

3 5 スピーカボタン

【手続補正27】

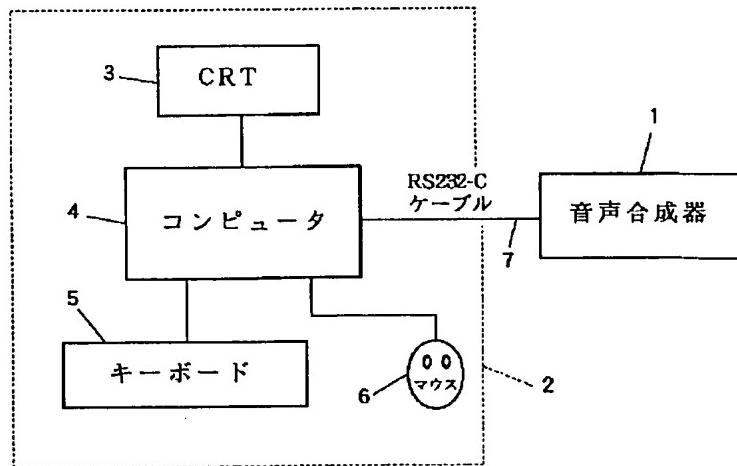
【補正対象書類名】図面

【補正対象項目名】図1

【補正方法】変更

【補正内容】

【図1】



【手続補正28】

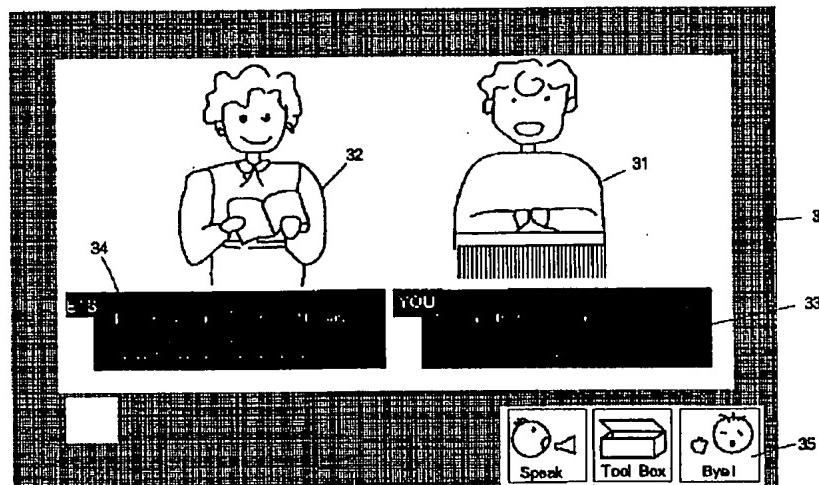
【補正対象書類名】図面

【補正対象項目名】図4

\* 【補正方法】変更

【補正内容】

\* 【図4】



【手続補正29】

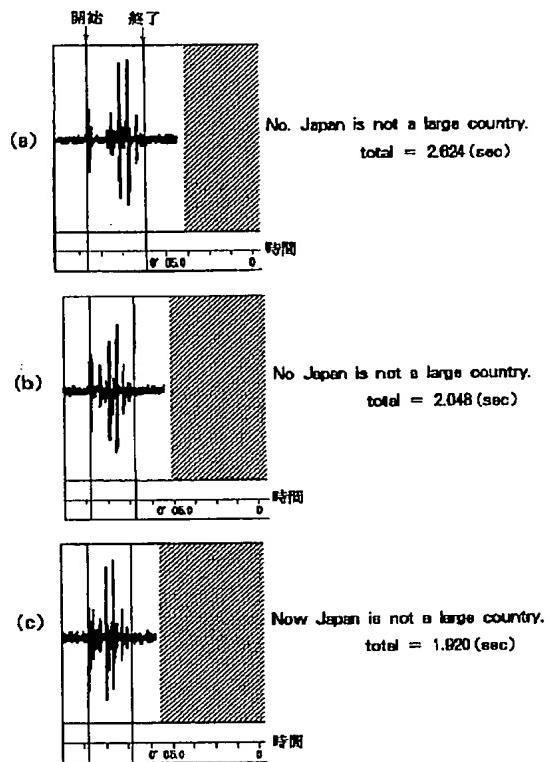
【補正対象書類名】図面

【補正対象項目名】図10

【補正方法】変更

【補正内容】

【図10】





JP5313686

[Biblio](#) | [Page](#) | [Drawing](#)**esp@cenet**

## DISPLAY CONTROLLER

Patent Number: JP5313686  
Publication date: 1993-11-26  
Inventor(s): SAKURAGI KEIKO; others: 03  
Applicant(s): SONY CORP  
Requested Patent:  [JP5313686](#)  
Application Number: JP19920109357 19920402  
Priority Number(s):  
IPC Classification: G10L3/00  
EC Classification:  
Equivalents:

### Abstract

PURPOSE: To output an optional voice and an image synchronously.  
CONSTITUTION: A voice synthesis part 12 calculates parameters for synthesizing a voice corresponding to a text supplied from a application program 14 through a control part 11. Then, when a mouse is operated and a speak button is clicked, a control part 11 outputs an actuation instruction to a voice synthesis part 12 and an animation display part 13, and the voice synthesis part 12 outputs the parameters for synthesizing the voice corresponding to the text to a voice synthesizer, which synthesizes the voice according to the parameters. At the same time, the animation display part 13 outputs an animation corresponding to the motion of the mouse for vocalizing the text in order.

Data supplied from the **esp@cenet** database - I2